WORKING PAPER
DEPARTMENT
OF ECONOMICS

QUASI-HYPERBOLIC DISCOUNTING AND RETIREMENT

Peter Diamond
Botond Koszegi

No. 00-03 January 2000

MASSACHUSETTS
INSTITUTE OF
TECHNOLOGY

50 MEMORIAL DRIVE
CAMBRIDGE, MASS. 02142

118

Quasi-Hyperbolic Discounting and Retirement

Peter Diamond and Botond Kőszegi[1]

December 16, 1999

[1]We are grateful to the National Science Foundation for research support (under grant SBR-9618698).

Abstract

There is overwhelming psychological evidence that some people run into self-control problems regularly, yet the effect of these findings on major life-cycle decisions hasn't been studied in detail. This paper extends Laibson's quasi-hyperbolic discounting savings model, in which each intertemporal self realizes that her time discount structure will lead to preference changes, and thus plays a game with her future selves. By making retirement endogenous, savings affect both consumption and work in the future. From earlier selves' points of view, the deciding self tends to retire too early, so it is possible that the self before saves less to induce her to work. However, still earlier selves think the pre-retirement self may do this too much, leading to possible higher saving on their part and eventual early retirement. Thus, the consumption path exhibits observational non-equivalence with exponential discounting. Observational non-equivalence also obtains on a number of comparative statics questions. For example, a self could have a negative marginal propensity to consume out of changes in future income. The outcome with naive agents, who fail to realize their self-control problem, is also briefly discussed. In that case, the deciding self's potential decision to retire despite earlier selves' plans results in a downward updating of available lifetime resources, and an empirically observed downward jump in the consumption path.

1 Introduction

If you are one of the vast majority of people who think they are saving *too* little of their income for retirement[1], the natural conclusion is that you have self-control problems. If, in addition, you argued to yourself that saving more today would only lead to spending more tomorrow, and thus there is no point in saving for retirement, at least there is a small consolation: you are a *sophisticated* decisionmaker with self-control problems. And self-control problems can extend beyond savings decisions. A thirty-something Italian one of us met in Prague, had decided that it wasn't worth looking for a job anymore, because even if he got himself to do it and found one, he would quit shortly thereafter, anyway.

It is exactly these kinds of agents our paper is mostly concerned with: people who have self-control problems but realize this and behave according to it. A very clean way to model such actors is through the introduction of quasi-hyperbolic discounting[2]. This form of discounting sets up a conflict between the preferences of different intertemporal selves, and thus introduces a need for self-control. With assumptions of no commitment and that the agent takes into account her self-control problem, savings decisions can then be modeled as an equilibrium in a sequential game played by the different selves. This modeling paradigm avoids the common connection made between preference changes and cognitive failures[3], and is therefore closer to standard economic analysis. The agent in the model understands perfectly the consequences of her actions, and acts optimally within the constraints imposed by her discount function, which the psychological evidence seems to support at least some of the time[4], and the absence of easily available commitment[5].

Laibson [6] analyzed actors of the above kind in detail. His key result is that sophisticated actors with a quasi-hyperbolic discount structure undersave; that is, all intertemporal selves could be made better off if all of them saved a little bit more. The reason for undersaving is that from the point of view of self t, self $t+1$ consumes too much, in essence wasting part of the savings inherited from self t. Self t doesn't like that, and she can do nothing about it, so she just gives self $t+1$ less. Both would love to commit self $t+1$ to saving more (self t because her savings wouldn't be wasted, and self $t+1$ because she would receive more savings), but the technology is simply not

[1]Bernheim [4] reports that people 'admit to' saving much less for retirement than they should. We don't know, though, how prevalent this is among academics.

[2]Quasi-hyperbolic instead of psychologically more accurate hyperbolic discounting is used only for computational tractability.

[3]For example, Mischel and Staub [8] find that subjects fail to understand the contingencies involved in a decision about delay of gratification. See Ainslie and Haslam [2] for further references.

[4]E.g., Ainslie [1].

[5]Admittedly, we are putting a somewhat unconventional twist on the discounted utility model. The DU model can only be justified if we put the agent's decision process entirely on the cognitive plane: unless we define utility in Baron's sense as the fulfillment of goals [3], which makes the definition of an instantaneous utility function extremely problematic, we can't say the decisionmaker experiences discounted utility in any way. The consequence of hyperbolic or quasi-hyperbolic discounting, namely, a conflict in the cognitive plane–where information is processed and decisions are made–then seems to imply irrationality on the part of the agent. This, however, is only so if we believe the understanding of humans as possessing an unchanging, non-contradictory essence. If the essence is changing or contradictory (that being outside our realm of choice and therefore not a question of rationality), that will translate into conflicts for our decision-making part. And in as much as self-control problems are exactly about the non-existence of such an entity, we definitely don't want to assume it for a study of them.

available.

We adapt Laibson's basic setup for the analysis of the effect of endogenous retirement decisions on savings behavior. The addition will be simply that in each of the models there will be a single period in which the agent can choose whether to work or retire. Working will cost the agent some utility, but she will be compensated with extra wealth. Commitment will not be possible: agents cannot precommit to a decision concerning retirement, nor to any consumption level.

We are mostly interested in building intuition through examining how the interaction with the retirement decision changes the Laibson savings problem. There are various benchmarks that involve the Laibson-type dynamics and that arise naturally in at least one of our models. All of them involve taking away the choice concerning retirement in one form or another. First, we can contrast equilibrium savings levels to those that would arise if there was a prior commitment possibility regarding work (delayed retirement), but no possibility of commitment about savings. At other times, a more useful comparison will be the simple Laibson-type equilibrium that results when the transition to retirement is determined exogenously, as, for example, with a mandate. The natural question to start with is whether a lack of commitment (or choice) results in higher or lower savings levels for retirement.

We will start with the simplest model that is relevant in (quasi-)hyperbolic discounting: the three-period model in which the middle period is the retirement decision period. Self 2 values the payoff compared to the effort she has to exert much less than self 1, so she will sometimes retire when self 1 would like her to work. In order to avoid this outcome, self 1 might save less (than she would if she could commit self 2 to work) to make self 2 work. On the other hand, if self 1 would like self 2 to work, but it is too expensive to achieve that without commitment, she will save more to help finance self 2's unavoidable retirement (than if she could commit self 2 to work). Note the qualitative distinction between a change in self 1's saving (compared to a setting with commitment) to induce a retirement decision and to accommodate one. Here, we can get lower saving to block the 'threat' of retirement and higher saving to accommodate it.

These effects are unchanged qualitatively by the addition of labor effort uncertainty or a longer horizon after retirement. Things get much more complicated when we allow for more periods before retirement. If lower saving is the outcome, it will be split by the pre-retirement selves–since the Laibson equilibrium is already characterized by overconsumption, earlier selves don't like to leave the job of undersaving to later selves. However, at least two periods before retirement new effects can also come into play. In particular, there can be a conflict between earlier and later selves about too *late*, not too early, retirement. This is because with quasi-hyperbolic discounting successive selves agree in what the later selves should do, they just don't agree on how much it is worth to induce them to do it. And the earlier self will always prefer higher savings levels than the later self. Thus we can get higher saving to 'encourage' early retirement. This effect exhibits much more mixed global implications than the lower saving one, mostly because earlier selves don't necessarily want to take part in the higher saving.

In the presence of a retirement decision there are a number of ways to observationally distinguish quasi-hyperbolic and exponential discounting. As Laibson has noted, in the savings game the path of consumption can't be used to distinguish the two, only some comparative statics observations can be [6]. This is not true if there is a retirement choice parameter: if higher saving or lower saving

(which are not possible with exponential discounting) happens, the consumption path won't look like that of an exponential discounter. In addition, there are numerous changes in the economy to which an agent endowed with quasi-hyperbolic discounting will react differently. The most radical diversion from the predictions of consistent preference models emerges when we consider the effect of an increase in wage level in the endogenous retirement period. If the agent undersaves to make the deciding self work, and the need for lower savings to induce work is relaxed through higher earnings, she will save more, giving a negative marginal propensity to consume out of changes in future income.

We realize, however, that many, if not most, of us fail to grasp fully the extent of our self-control problem. That is, we often fail to be fully sophisticated in the sense of the Laibson model. Therefore, we will briefly discuss the potential outcomes under the particular assumption of naiveté, that each self falsely assumes that the others will comply with her plans. Since there is no game in this case, the analysis is considerably simpler. One interesting implication of naiveté is the possibility that the selves before the deciding self plan to retire late, but the deciding self chooses to retire early, leading to an update in lifetime wealth and thus a drop in the consumption path at retirement.

2 The quasi-hyperbolic discounting setup

We adapt the structure recently used by Laibson for analyzing quasi-hyperbolic discounting issues. For a more detailed introduction, see [6], for example. The consumer's instantaneous utility function is of the constant relative risk aversion (CRRA) class, that is

$$u(c) = \frac{c^{1-\rho}}{1-\rho} \text{ if } \rho \neq 1, \text{ and } u(c) = \ln(c) \text{ if } \rho = 1, \tag{1}$$

ρ being the risk aversion parameter. A nice property of CRRA utility functions is the fact that for intertemporal maximizations of the form

$$\begin{aligned} &\max_{c_1, c_2} u(c_1) + \kappa u(c_2) \\ &\text{s.t. } c_1 + \frac{1}{R} c_2 = W \end{aligned} \tag{2}$$

with κ a positive discount factor, the solution will always be $c_1 = \lambda(R, \kappa)W$ for some $0 < \lambda(R, \kappa) < 1$. Also, then, some easy manipulation shows that lifetime discounted utility can be written as $K(R, \kappa)u(W)$ (or $K(R, \kappa) + u(W)$ for $u(c) = \ln(c)$) for a positive function $K(R, \kappa)$. This allows us to collapse periods where we have already solved the problem and gotten linear answers into a single period, a shortcut extremely convenient for backward induction arguments. We will use this property a number of times in the paper.

In a T-horizon game, self t's discounted utility from present and future consumption is

$$u(c_t) + \beta \sum_{i=1}^{T-t} \delta^i u(c_{t+i}) \tag{3}$$

with an expectation at front if there is uncertainty. β and δ (both between 0 and 1) are discount parameters meant to capture the essence of hyperbolic discounting, namely that the discount factor between adjacent periods close by is smaller than between similar periods further away. Indeed, the discount factor between periods t and $t+1$ is $\beta\delta$, and between any two adjacent periods later it is δ.

Of course, the discount structure just described applies only to self t; for example, self $t+1$'s discount factor between $t+1$ and $t+2$ is $\beta\delta$. Therefore, there is a conflict between different selves about how much to consume (or whether to retire) in a given period, or, more formally, preferences are intertemporally inconsistent. We assume that commitment is not possible (so that each self controls her period's consumption, subject to a financial or wealth constraint[6], and possibly a decision concerning retirement), and model the behavioral decisions as a subgame-perfect equilibrium of the game played by the different selves[7]. Finally, R is the constant and exogenous gross return on wealth.

3 The three-period model

We begin with the three-period model, the shortest possible that actually generates quasi-hyperbolic discounting effects. The periods are labeled 1,2,3, and subscripts on c or W refer to the period in question. In the first period, the agent has to work; in the second, she can decide whether to work or retire; and in the third, she has to retire[8]. The agent incurs a constant utility cost of effort $e > 0$ if she works in the second period, but she also gets an extra Δ amount of income if she does.

As usual when looking for subgame-perfect equilibria, we solve backwards. The decision is easy in the third period: no work is done and all remaining wealth is consumed. Suppose, then, that the period 2 self inherits a wealth of W_2. This will be her remaining wealth if she retires, and she will have $W_2 + \Delta$ if she works. As we have mentioned above, there is a $\lambda > 0$ such that self 2 will always consume a proportion λ of her wealth. Thus her discounted utility is

$$u(\lambda W_2) + \beta\delta u(R(1-\lambda)W_2) \tag{4}$$

if she doesn't work, and

$$u(\lambda(W_2+\Delta)) + \beta\delta u(R(1-\lambda)(W_2+\Delta)) - e \tag{5}$$

if she works. Therefore she will work iff[9]

$$u(\lambda(W_2+\Delta)) - u(\lambda W_2) + \beta\delta u(R(1-\lambda)(W_2+\Delta)) - \beta\delta u(R(1-\lambda)W_2) \geq e. \tag{6}$$

[6]depending on whether there are liquidity constraints.

[7]The game theory-based decision rule is basically equivalent to the assumption of sophistication on the part of the agent. An alternative assumption is naiveté, where each self naively assumes that others will follow her decisions. We will study naifs briefly in section 7.

[8]Making the key period–that of the retirement decision–the first or third period takes all the spice out of the model because then the conflict to be described below wouldn't materialize.

[9]We are assuming for now that the agent will work if she is indifferent. In the long-horizon models, we will more generally assume that an agent indifferent between two actions will choose the one the earlier selves would prefer. (With quasi-hyperbolic discounting, all earlier selves want the same thing.) It turns out that this gives the essentially unique subgame-perfect equilibrium–otherwise, the earlier self's maximization problem has no solution.

Since u is concave, there is a $\overline{W}_2$ such that self 2 will retire iff $W_2 > \overline{W}_2$.

Now let's look at this situation from the point of view of self 1. She will prefer self 2 to work if

$$\beta\delta u(\lambda(W_2+\Delta)) - \beta\delta u(\lambda W_2) +$$
$$+\beta\delta^2 u(R(1-\lambda)(W_2+\Delta)) - \beta\delta^2 u(R(1-\lambda)W_2) \geq \beta\delta e,$$

or

$$u(\lambda(W_2+\Delta)) - u(\lambda W_2) + \delta u(R(1-\lambda)(W_2+\Delta)) - \delta u(R(1-\lambda)W_2) \geq e. \tag{7}$$

Notice that the left-hand side of 7 is greater than the left-hand side of 6; consequently, there is a range of wealth levels for which self 2 won't work, but self 1 would like her to. In particular, for $W_2 = \overline{W}_2$ self 2 is indifferent between working and not working, but self 1 strictly prefers her to work. This effect arises simply because self 1 weighs the cost and the benefit of working differently: for her, the cost is less salient.

Figure 1 displays the continuation utility for self 1 (her utility from periods 2 and 3) as a function of W_2, the level of wealth self 1 leaves for self 2, for an example with logarithmic utility. The curve that starts off as a solid line and continues as a dotted one (U_w) is self 1's utility *assuming* self 2 works, and the other curve (U_r) is her utility assuming self 2 doesn't work. Only the solid part of each curve is available to self 1, as she has to take into account self 2's decision. Nevertheless, the simplest way to understand self 1's maximization problem is through U_w and U_r. Define s_i^* (i being r or w) to be the wealth received by self 2 in the solution to the maximization problem that ignores the endogeneity of retirement:

$$\max_s u(W_1 - \frac{1}{R}s) + U_i(s). \tag{8}$$

Note that $s_w^* < s_r^*$ since work provides extra income in period 2, and some of that is consumed in period 1. If self 1 could commit self 2 to a decision on work (but not on consumption), she would choose one of these savings levels. Therefore, let s_c^* be the better of the two possibilities s_i^*. That is, s_c^* gives the optimal savings level if self 1 could commit self 2's retirement decision. We describe optimal savings levels by examining these constructs.

If self 1 would commit self 2 to retire, that is, $s_c^* = s_r^*$, then $s_r^* > \overline{W}_2$, the level of wealth at which self 2 is indifferent to retirement. Otherwise s_r^* would be dominated by a working alternative. Consequently, in that case the optimal savings level is $s^* = s_c^* = s_r^*$; an inability to commit to retirement does not matter if self 1 wants self 2 to retire. If $s_c^* = s_w^*$, things take a more interesting turn. If $s_w^* \leq \overline{W}_2$, then analogously to the above $s^* = s_c^* = s_w^*$. However, it is also possible that $s_w^* > \overline{W}_2$, so that s_c^* and work in period 2 is not available to self 1. Then–since U_w is concave–the best point on the available part of U_w is the boundary, $s_b^* = \overline{W}_2$. This might or might not dominate the best point resulting in retirement, which satisfies $s_r^* > s_w^* > \overline{W}_2$. Thus, two possibilities emerge: either $s^* = s_b^* = \overline{W}_2 < s_w^* = s_c^*$, or $s^* = s_r^* > s_w^* = s_c^*$. Notice that in each of these cases the equilibrium savings level is different from the one under commitment.

Figure 2 shows lifetime discounted utilities for self 1 as a function of W_2 assuming work and retirement in period 2 for the same example as in figure 1. Again, the solid part of each curve is available to self 1. s_w^* maximizes the work curve, s_r^* the retirement curve, and, as we have seen,

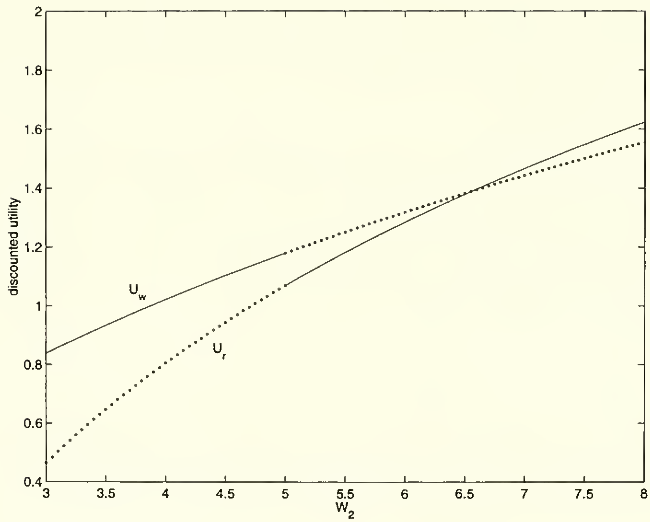

Figure 1: Utility of self 1 from periods 2 and 3 with and without work in period 2

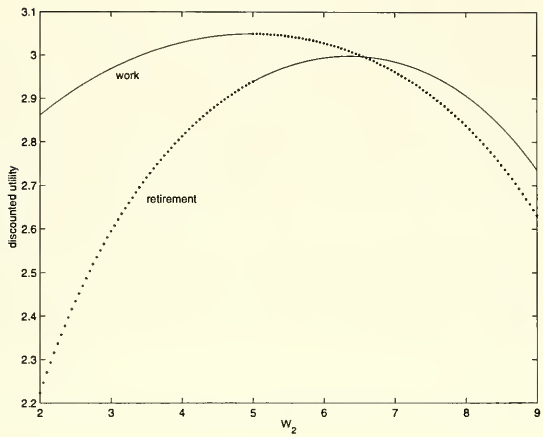

Figure 2: Lifetime utility of self 1

the best available point among s_w^*, s_r^*, and $\overline{W}_2$ is the optimum for self 1. In this example, it seems to be $\overline{W}_2$.

We can trace out the equilibrium as a function of W_1 for the general case. For low values of W_1, $s_w^* < \overline{W}_2$, self 2 works in equilibrium and the inability of self 1 to commit self 2 to work has no effect. Then, there is a range of values for W_1 such that optimal savings equals $\overline{W}_2$ in order to just induce work. Over this range savings are less than they would be if self 1 could commit self 2 to work. In the next range of W_1, self 1 accommodates self 2, saving $s_r^* > \overline{W}_2$ even though self 1 would save less and commit self 2 to work if that were possible. For high enough values of W_1, self 1 prefers that self 2 retire and there is, again, no effect from the inability to commit. This is shown in figure 3.

The marginal propensity to consume in period 1 out of a small increase in W_1 behaves differently in the different regions[10]. In the lowest region, for a small increase in W_1, the fraction of the increase consumed is λ_1, just as in the case without a retirement decision. For a small increase in wealth in the second region, all of it is consumed so that self 2 continues to receive $\overline{W}_2$. For small increases in wealth in the top two regions, again, the fraction λ_1 is consumed in period one.

When interpreting results in these short-horizon models, we have to be very careful not to confuse genuine quasi-hyperbolic discounting effects with effects that arise due to the fact that we have chosen a short horizon. In particular, you might notice that even if $s^* = s_i^*$ for an i, we don't have $\frac{c_1}{c_2} = \frac{c_2}{c_3}$ as we do for exponential discounting with CRRA utility functions. But this peculiarity occurs only because the marginal propensities to consume change from period to period,

[10]By a small increase we mean one that does not move self 1 into a different region.

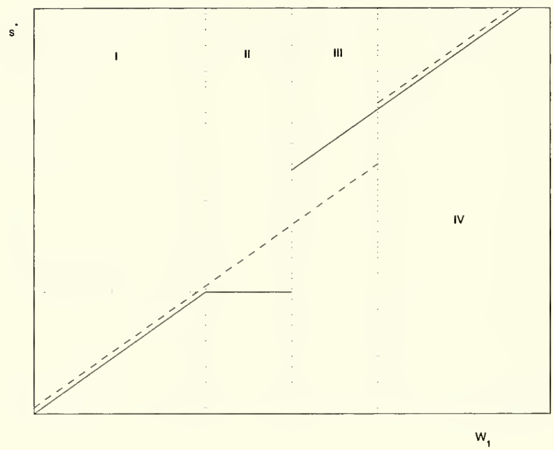

Figure 3: Savings of self 1 with (dashed line) and without (continuous line) commitment
Note: the figure is only qualitative; it is not meant to illustrate actual slopes or relative sizes for the regions.

a property that disappears as the horizon after retirement is assumed to go to infinity[11]. In that case, only the equivalent of $s^* = \overline{W}_2$ will not satisfy the equivalent of $\frac{c_1}{c_2} = \frac{c_2}{c_3}$. But you don't have to turn to section 5 to understand that the case $s^* = \overline{W}_2$ is the only one observationally different from exponential discounting: it is the only case when self 1 uses non-optimal savings (in the sense of the consumption game) as a tool to change the retirement decision of self 2. And as Laibson has pointed out in the context without a retirement decision, optimal savings with quasi-hyperbolic discounting is observationally equivalent to exponential discounting [6]. Non-optimal savings, finally, is not possible with exponential discounting, even in the presence of a retirement decision: in that case $\overline{W}_2$ is defined by the intersection of the curves U_w and U_r, so s_w^* and s_r^* both dominate it, and s_c^*, which is always available, does so strictly.

Behaviorally, as opposed to just observationally, there is another, more subtle, difference between quasi-hyperbolic and exponential discounting. It is possible that self 1 would prefer to commit self 2 to work and give her $s_c^* = s_w^*$, but since that is not possible and she has to undersave too much to make self 2 work, she chooses $s^* = s_r^*$. This reason for choosing s_r^*, though unobservable, *is* unique to quasi-hyperbolic discounting: it arises from the conflict of self 2's decisions and self 1's

[11] The marginal propensity to consume matters with quasi-hyperbolic discounting simply because the Euler equation contains it:

$$\frac{u'(c_t)}{u'(c_{t+1})} = R\delta\left(\beta\frac{\partial c_{t+1}}{\partial W_{t+1}} + 1 - \frac{\partial c_{t+1}}{\partial W_{t+1}}\right).$$

This is proved in [6] but also falls out as a special case of our analysis in section D of the appendix.

wishes. However, the reason for higher saving in this case (clearly $s^* > s_c^* = s_w^*$) is very different from the reason for lower saving above[12]: it is not intended to change the retirement decision of self 2. Quite the opposite: in recognition of the fact that it would be 'too expensive' to change self 2's decision, self 1 will save more to offset the lower wealth level of self 2 due to the early retirement.

This is a point where the addition of a retirement decision fundamentally changes the Laibson problem. With only savings, the 'wasting' of wealth on the part of later selves always decreases the marginal utility of savings for the earlier self (that's why that self saves less). When there is a retirement decision, however, a 'wrong decision' by the later self can decrease wealth, and thus increase the marginal utility of savings[13].

Though very simple, the three-period certainty model delivers much of the intuition that we will encounter in more complicated models. In the next section, we show that the different cases discussed here are robust to the addition of uncertainty. Adding uncertainty will eliminate case analysis and transform it into effects analysis, thereby also allowing a delineation of when higher or lower saving is likely to occur.

4 Uncertainty

We could introduce uncertainty in period 2 labor income (Δ), and in period 2 cost of effort (e). The two give similar results, and the latter is somewhat nicer to present, so we present only that one. Assume therefore that self 1 doesn't know e, but knows its continuous distribution function f (and the cumulative distribution function F). This standard assumption is made plausible by the possibility that the agent does not know how healthy or how thrilled she will be to work in the future. We assume that the support of f is wide enough to encompass all of the regions above.

We start again from self 2's problem, who has inherited a wealth W_2. Define $\bar{e}(W_2)$ as the level of effort cost at which self 2 is indifferent to work:

$$u(\lambda(W_2+\Delta)) + \beta\delta u(R(1-\lambda)(W_2+\Delta)) - \bar{e}(W_2) = u(\lambda W_2) + \beta\delta u(R(1-\lambda)W_2) \qquad (9)$$

Self 2 will work if $e \leq \bar{e}(W_2)$. Therefore self 2 will work with probability $F(\bar{e}(W_2))$ and retire with probability $1 - F(\bar{e}(W_2))$. For simplicity, let K be the constant such that $\beta\delta u(\lambda W) + \beta\delta^2 u(R(1-\lambda)W) = Ku(W)$. As we have mentioned, such a constant always exists for CRRA utility functions[14]. Now the maximand for self 1 is

$$u(W_1 - \frac{1}{R}W_2) + K[F(\bar{e}(W_2))u(W_2+\Delta) + (1-F(\bar{e}(W_2)))u(W_2)] - \beta\delta\int_0^{\bar{e}(W_2)} ef(e)de. \qquad (10)$$

The first-order condition is

$$\frac{1}{R}u'(W_1 - \frac{1}{R}W_2) = K[F(\bar{e}(W_2))u'(W_2+\Delta) + (1-F(\bar{e}(W_2)))u'(W_2) +$$

[12] Remember that the benchmark for all savings discussions at this point is the savings level that would arise if self 1 could control self 2's retirement decision.

[13] The $s^* = \overline{W}_2$ case has an interpretation in the Laibson spirit: since savings would be wasted by self 2 to finance retirement, self 1 saves less to not let self 2 do that.

[14] When the utility function is logarithmic, the correct expression is $\beta\delta u(\lambda W) + \beta\delta^2 u(R(1-\lambda)W) = K + u(W)$. The analysis is the same in this case.

$$f(\overline{e}(W_2))\overline{e}'(W_2)u(W_2+\Delta) - f(\overline{e}(W_2))\overline{e}'(W_2)u(W_2)] - \beta\delta\overline{e}(W_2)f(\overline{e}(W_2))\overline{e}'(W_2),$$

which is equivalent to

$$\frac{1}{R}u'(W_1 - \frac{1}{R}W_2) = K[F(\overline{e}(W_2))u'(W_2+\Delta) + (1-F(\overline{e}(W_2)))u'(W_2)] + \\ f(\overline{e}(W_2))\overline{e}'(W_2)[Ku(W_2+\Delta) - \beta\delta\overline{e}(W_2) - Ku(W_2)]. \quad (11)$$

A similar first-order condition would arise if self 1 could commit self 2 to a state-contingent retirement decision[15], except that $\overline{e}(W_2)$ should be replaced by $\tilde{e}(W_2)$, where $\tilde{e}(W_2)$ is defined by

$$\beta\delta u(\lambda(W_2+\Delta)) + \beta\delta^2 u(R(1-\lambda)(W_2+\Delta)) - \beta\delta\tilde{e}(W_2) = \beta\delta u(\lambda W_2) + \beta\delta^2 u(R(1-\lambda)W_2). \quad (12)$$

(This just defines the cutoff cost level under which self 1 would want self 2 to work.) Then, by definition, $Ku(W_2+\Delta) - \beta\delta\tilde{e}(W_2) - Ku(W_2) = 0$, so the first-order condition is

$$\frac{1}{R}u'(W_1 - \frac{1}{R}W_2) = K[F(\tilde{e}(W_2))u'(W_2+\Delta) + (1-F(\tilde{e}(W_2)))u'(W_2)]. \quad (13)$$

Neither of these two first-order conditions is well-behaved, and we have not found simple conditions on f that would make them well-behaved. What we would like is for the right-hand sides of equations 11 and 13 to be decreasing in W_2. Then we would have unique solutions to the first-order conditions, which would be global maxima. Notice that the derivative of the right-hand side of 11 is of the form

$$K[F(\overline{e}(W_2))u''(W_2+\Delta) + (1-F(\overline{e}(W_2)))u''(W_2)] + f(\overline{e}(W_2))[Z] + \\ f'(\overline{e}(W_2))\overline{e}'^2(W_2)[Ku(W_2+\Delta) - \beta\delta\overline{e}(W_2) - Ku(W_2)], \quad (14)$$

where the expression Z multiplied by $f(\overline{e}(W_2)$ is complex and not worth writing down for our purposes. The derivative of the right-hand side of 13 is very similar, the difference being that $\overline{e}(W_2)$ is replaced by $\tilde{e}(W_2)$ and there is no term multiplied by f'[16]. Now it is easy to see that if f and f' are 'small enough' (though it is hard to give meaning to this phrase), the problem is well-behaved. For example, a uniform distribution with a large enough support will do. This is certainly a sufficient condition, albeit not necessary.

Having said that, we assume that unique solutions to the FOCs exist, in which case they define the maximum. We are interested in the difference of the right-hand-sides of the first-order conditions[17]:

$$K\overbrace{[F(\tilde{e}(W_2)) - F(\overline{e}(W_2))][u'(W_2) - u'(W_2+\Delta)]}^{\text{higher saving}} +$$

[15] A commitment device conditional on the realized e is not very realistic, but as a comparison it is useful for highlighting the tradeoffs self 1 faces. If self 1 could only commit to a specific decision (one not conditional on e), she would never commit to retirement, and to work only if that is not too costly on the high-e end.

[16] The term multiplied by $f(\overline{e}(W_2)$ is also simpler.

[17] If the difference is positive at the optimal savings with commitment, then the optimal savings without commitment is higher. This is trivial if the problem is well-behaved in the above sense. But the assumption that the first-order condition has a unique solution, together with the observation that for low W_2 the right-hand side of equation 11 is greater than the left-hand-side, and vice versa if W_2 is close to RW_1, is also sufficient. Similarly, the opposite is the case if the difference is negative.

$$f(\overline{e}(W_2))\overline{e}'(W_2)\underbrace{[Ku(W_2+\Delta)-\beta\delta\overline{e}(W_2)-Ku(W_2)]}_{\text{lower saving}} \tag{15}$$

Notice that since $\tilde{e}(W_2) > \overline{e}(W_2)$ for any W_2, the overbraced product is positive, so it indeed encourages higher saving. On the other hand, we know that for $\overline{e}(W_2)$, self 2 is indifferent between working and not working, and also that in that case self 1 would prefer her to work. Thus the underbraced term is positive. But $\overline{e}'(W_2)$ is negative, so the given effect in fact tends to lower savings.

The intuition behind these two effects is straight-forward enough. First, since there is a chance that self 2 will retire when self 1 prefers that she work, she'll need more money than if she worked. Thus, self 1 saves more. Second, since saving less induces work in some additional states, self 1 has an incentive to save less.

It should be clear that these are just translations of the cases analyzed in the certainty model into the uncertainty setting. This setup, in addition, also allows for convenient analysis of when higher or lower saving is likely to occur. For example, if $f(\overline{e}(W_2^c))$ (where W_2^c is optimal savings with commitment) is high compared to $F(\tilde{e}(W_2^c)) - F(\overline{e}(W_2^c))$, we will get lower saving. That is, if self 1 feels that she can exert a lot of influence on self 2's decision through savings, she will save less. On the other hand, if $f(\overline{e}(W_2^c))$ is close to zero, while $F(\tilde{e}(W_2^c)) - F(\overline{e}(W_2^c))$ is fairly large, there will be higher saving. In simpler terms, if self 1 can't exert much influence on self 2, she will just accept that self 2 might retire too early, and give the now poorer self more savings[18].

5 Multiple periods before the retirement decision

In the savings game, the tendency of long horizons to make marginal propensities to consume approximately equal across periods helps both in describing the quasi-hyperbolic equilibrium and in comparing it with the exponential discounting outcome. Put differently, a long but finite horizon is a convenient tool to pin down the equilibrium while keeping the smoothness properties of an infinite horizon. But it doesn't offer many surprises. This is also the case in our model if the horizon after retirement is long. To create a 'disagreement' between the self making the work/retirement decision and the previous one, which is what drives the results of our previous sections, one only needs a setting in which the cost of extra work is concentrated in a single period, while the benefits of it are spread out. This is certainly satisfied if we have a long horizon after retirement, so as long as there is only a single period before the decision period, we shouldn't expect the results to change too much. In fact, they don't; but since a detailed discussion would add little, it is relegated to the appendix.

Though introducing a long horizon after retirement is of little consequence to the qualitative results of our models, a longer horizon before retirement does set up a novel, interesting distinction: how the effects play themselves out close to versus far from retirement[19]. Unlike in the savings

[18]Notice that making the size assumptions on f and f' does not make the comparison of the two effects an irrelevant exercise. Though f and f' are small (compared to 1), there is no restriction on their *relative* size, so $f(\overline{e}(W_2^c))$ and $F(\tilde{e}(W_2^c)) - F(\overline{e}(W_2^c))$ might compare in any number of ways.

[19]As we have seen in the previous section, uncertainty helps in putting effects nicely side-by-side. Also, it seems

game, what happens at the end is no longer an empirically unattractive theoretical nuisance; the behavior is not a response to nearby deterministic death, but to the approach of the end of working life–the central focus of this paper. Thus, we will 'move backwards' in this section, and see what happens when the horizon before retirement is let to grow. Perhaps surprisingly, the effects change considerably.

Unfortunately, there is very little we can say about the equilibrium in general with many periods of work. The bulk of the trouble stems from the fact that when later selves have decreasing marginal propensities to consume[20], the consumption schedules become extremely complicated very quickly as we move to earlier periods.

All we know is that the agent's consumption schedule is piecewise linear in wealth for each t, and, furthermore, the agent's consumption path is as if she was going through a series of shorter Laibson problems[21]. This is quite interesting in itself: the agent periodically acts as if she is liquidity constrained and/or impatient, even though she has perfect foresight and faces no constraints. But since we are unable to say much in general about the equilibrium, we will mostly restrict our attention to a model in which there are only two periods of exogenously mandated work before the period of endogenous decision, though some results will be more general than that. For notational convenience, we now label the period with a retirement decision as period 0–the first period of life is thus period -2. We assume a long horizon after retirement, although the qualitative results are the same with a shorter lifespan[22].

Before we plunge into the work, we reconsider the benchmark for our savings discussions. For the three-period model, the difference from the benchmark case of what would happen if self 1 could commit self 2 to a retirement decision answered the question of how the conflict between selves 1 and 2 was reflected in savings. For longer horizons before retirement, the conflict is not only between the current self (say self t) and the self making the retirement decision–there are many selves in-between with whom self t may also have a conflict. An important implication of this is that self t *might not want to* commit to a retirement decision. Commitment also allows other selves to behave differently, which self t might not like[23].

Fortunately, this issue is not too critical if the earliest self considered is self -2, so we will keep using similar vocabulary to section 3, while realizing that this would be inappropriate for longer horizons.

As we have mentioned, self -1's behavior is qualitatively the same as what we have seen in

to make the problem smoother, eliminating discontinuities in selves' strategies [7]. However, there is only an Euler equation to work with, and for a long horizon before retirement we know little about the global properties of the solution. While this might be satisfactory in a savings problem, it is inherently unfitted for analyzing the retirement decision: the Euler equation says nothing about it. So we will have to do with the certainty model for this section.

[20]Laibson [7] describes such an example in detail, though in the context of liquidity constraints. Here, since self 2 in section 3 has a region where her marginal propensity to consume is 1, in that region she behaves as if 'liquidity constrained.' This gives the jumps in consumption earlier on.

[21]See the appendix for a formal statement and a proof.

[22]Also, we will assume in this section that a Markov-perfect equilibrium in pure strategies exists for the game. The proof of this claim is contained in the appendix. Noticeably, that proof uses similar methods to those below, but putting it in the appendix and just assuming existence for now makes the paper much easier to follow.

[23]We could say that we are comparing things to when self t is forced to make a commitment, but if that is against self t's will, the interpretation of the results is ambiguous.

section 3. Let us now move back to self -2 and see what she thinks about the behavior of self -1. (Notationally, we include the PDV of earnings in all periods except 0 in the wealth measure W, considering only the possible earning Δ in period zero separately.) Suppose giving the next self the savings level in a Laibson problem ($W_{-1} = R(1-\lambda^*)W_{-2} - \frac{1}{R}\lambda^*\Delta$) would result in a savings decision $s^* < s_c^*$ by self -1. In the pure Laibson problem with work in period zero, self -1 is already consuming too much from the point of view of self -2 (even the savings level s_c^* is too low for her), so an even larger consumption should leave self -2 rather unhappy. If self -2 likes the working alternative, she would rather do part of the 'overconsumption' herself. And we know she can do that, since for lower wealth levels self -1 still prefers late retirement. In fact, we have the following more general, intuitive result:

Lemma 1 *Suppose that $t \leq 0$ and $W_t > W_t'$. Then it is not possible that self t with wealth W_t behaves so that self 0 eventually works, and with wealth W_t' she behaves so that self 0 eventually retires.*

The formal proof is in appendix B. It takes advantage of the concavity of consumption utility to show that savings is monotonically increasing in wealth for each self before zero. This implies that self 0's wealth is monotonically related to previous selves' wealth levels. And we know self 0 retires iff $W_0 > \overline{W}_0$ for a given $\overline{W}_0$.

The above result rested on the assumption that self -2 liked the working alternative. What if she doesn't? We have already seen that she views the tradeoff between the extra wealth and making self 0 work differently from self -1. But even without lower saving by self -1, she certainly views the benefit and cost of working an extra period differently. If so, what kind of conflicts arise between self -2 and self -1?

The answer might be surprising: self -2 will never use boundary (knife-edge) savings to get the working alternative, but it is possible she will use it to 'force' retirement. This is exactly what the following lemma proves.

Lemma 2 *Let $\overline{W}_t$ ($t < 0$) be the level of wealth at which self t is indifferent between behaviors that eventually lead to self 0 working or retiring. At this savings level, self $t-1$ strictly prefers self t to choose to eventually make self 0 retire.*

Once again, the proof is in the appendix, but it's essence is simple: due to the different preferences, self t cares relatively more about consumption in period t than does self $t-1$, so when self t is indifferent, self $t-1$ wants her to go for the low-consumption (high-saving) option. And this is of course the early retirement option. Self -2, then, might save more than with mandated early retirement to just induce self -1 to save so as to result in early retirement.

These lemmas can be used to illustrate self -2's general qualitative savings behavior relative to wealth, which is done in figure 4. For very low levels of wealth, self -2 prefers late retirement, and this can be achieved under the Laibson consumption solution.

In the next two regions (II and III), selves -2 and -1 undersave to induce self 0 to work. By lemma 1, self -2 can split the undersaving with self -1, while still inducing eventual late retirement. For relatively low wealth levels where there has to be undersaving done to induce self 0 to work

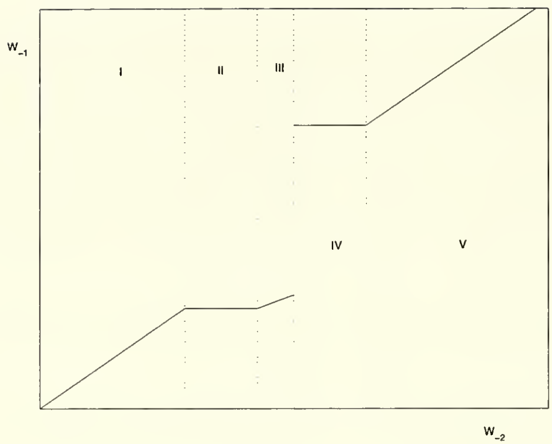

Figure 4: Savings of self -2

Note: the figure is only qualitative; it is not meant to illustrate actual slopes or relative sizes for the regions.

(region II), all the undersaving will be done by self -2. In this region, self -1's marginal propensity to consume is 1, so self -2 prefers to consume all extra wealth as long as $u'(c_{-2}) > \beta\delta u'(c_{-1})$, and her savings function is flat as a function of wealth. As self -2 gets richer, she will want to split the extra consumption with self -1 even though self -1 has a marginal propensity to consume of 1. In this region (III), we have $u'(c_{-2}) = \beta\delta u'(c_{-1})$, and the savings function is positively sloped, although with a lower slope than in region I [24].

Region IV is the content of lemma 2–self -2 chooses early retirement, but in order for her to do that, she needs to oversave, otherwise self -1 ends up making self 0 work. Again, in this region self -2 consumes all extra marginal wealth, until she is rich enough so that eventual early retirement results without oversaving. And in region V, self -2 consumes according to the Laibson solution,

[24]There are some things that can be said in greater generality. Assume that each self $t < 0$ already prefers early retirement for a low enough wealth level so that there are no jumps in self t's consumption function on the late retirement section. (We expect the statements that follow to be true even without this assumption, but haven't been able to prove it.) Then it is easy to prove by backward induction and taking advantage of the above lemma that two things are possible. Either self t's marginal propensity to consume is λ^* up to $(\frac{1}{(1-\lambda^*)R})^t \overline{W}_0$, and (if she still prefers late retirement for higher wealth levels) then her marginal propensity to consume is 1 on some non-empty interval. Or self t's marginal propensity to consume is λ^* up to some lower wealth level, above which she prefers early retirement. This immediately implies two things. First, if mandated work is acceptable to all selves (in the sense that they prefer late retirement at their mandate wealth level), then the outcome of a mandate is an equilibrium even with choice. Second, if this is not the case (the mandate is not acceptable to all selves), then savings for retirement in a work equilibrium without a mandate is lower. Also, if all selves prefer lower saving for at least some wealth levels, then small enough amounts of lower saving will all be done by the first self alive.

leading to early retirement.

More generally, for any wealth level W_t with $t \leq -2$ such that self t wants self 0 to retire, self $t-1$ wants her to retire as well. The converse of this is not true, i.e. if self t chooses to save so that self 0 works, self $t-1$ might not like that. Translating our intuition from the work equilibrium, we might be led to think that–due to the elimination of this conflict–if retirement was mandated, savings levels would be lower. Such a conclusion is true in the present setup, but not if we go back one more period. Imagine that with the mandate, W_{-1} is slightly above $\overline{W}_{-1}$, the cutoff wealth level for self -1, and that self -2 is willing to bequeath higher savings to make self -1 choose early retirement. That is, even for some wealth levels below $\frac{1}{R}\frac{1}{1-\lambda^*}\overline{W}_{-1}$, self -2 will choose to save $\overline{W}_{-1}$. Since self -2 overconsumes from the point of view of self -3, self -3 might choose to lower her savings to self -2 once the mandate is removed. Then self -1 will end up with $\overline{W}_{-1}$–lower than with the mandate. The key intuition is that self -3 takes advantage of self -2's efforts to control self -1's decision for her own purposes[25].

This highlights a key distinction between the early and late retirement outcomes. When there is higher saving to be done to force retirement, the earlier selves are by no means as eager to join in as when the task is lower saving. They are actually very happy to let later selves save more, as those selves consume too much from their point of view anyway. They will thus want to have them oversave a lot, often resulting in putting the self at her cutoff wealth level. (In more precise language: a self t who is leaving boundary saving but over her own cutoff wealth level usually has a marginal propensity to consume of 1, thus making the marginal rate of substitution for self t-1 low.) As a consequence, small amounts of lower savings are 'handled' by the early selves, while higher saving is pushed on (in an exaggerated manner, in fact) to later ones.

Partly for this reason, it is important to focus on the lower saving outcome if we care about the well-being of the individual as a whole, that is, the set of her intertemporal incarnations. For such an analysis we can use similar tools as in welfare economics. In the 'forced' work outcome, the implications are generally bad. The Laibson consumption path is already too high, and there is additional consumption done in the periods before retirement, making the equilibrium outcome Pareto-inferior to a mandate: self -2 would benefit from a better consumption path, selves 0 and up from more savings, and self -1 (possibly) from both. In this strong sense, the equilibrium outcome is suboptimal, and can correctly be termed an undersaving outcome. Similarly unambiguous things cannot be said when the equilibrium has retirement in period 0. Higher saving by a self is in general good for both earlier and later selves but bad for that self. So, on the one hand, commitment might not be desirable, and on the other, its welfare implications are mixed.

In all these proofs we have *very strongly* used the particular structure of quasi-hyperbolic discounting[26]. A troublesome occurrence of this was when we proved that in periods $t \leq -2$ lower saving is not possible in the boundary savings level sense (lemma 2): the proof depended on the fact that selves t and $t-1$ have two different weightings of the same utility tradeoff (c_t vs. K_t). Since Laibson introduced quasi-hyperbolic discounting as an approximation to hyperbolic discounting

[25]The same counterexample works to show that the other statement from the late retirement case does not carry over, either: it is not true that if mandated retirement is acceptable to all selves, then the outcome of the mandate is an equilibrium.

[26]Even the appendix's proof of the existence of equilibria uses at a crucial point that with quasi-hyperbolic discounting all earlier selves would want a later self to do the same thing.

purely for analytical convenience, such results should be handled with great suspicion. In addition, our intuition should revolt at strange answers of this sort. In a true hyperbolic discount structure, from the point of view of self -2, self -1 not only underweights effort in period 0 compared to consumption in period -1, but she also overweights it compared to consumption after retirement. This results in self -2 choosing to undersave more often than in a quasi-hyperbolic model, where the second conflict is nonexistent. For a formal treatment, see the appendix.

6 Notes on observational equivalence

One of the important caveats of quasi-hyperbolic discounting is that it is very hard to tell it apart from exponential discounting. Laibson [6] noted that an econometrician watching a quasi-hyperbolic discounter, but operating under the assumption of exponential discounting, will get a very good fit for her theory, as consumption paths of the two types of agents look exactly the same. At the same time, she will radically misconstrue the agent's preferences, finding a one-period discount factor of 0.98 instead of 0.6 in a typical example. Only comparative statics involving the interest rate can be used to distinguish actors with self-control problems from the others.

Our models lend themselves to a number of convenient approaches to this question. Both the consumption path and some comparative statics results can give a quasi-hyperbolic discounter away.

First, a consumption path that is smooth after retirement and not smooth leading up to it is a sign of quasi-hyperbolic discounting. This is of course due to the changing marginal propensities to consume in the periods preceding retirement. In particular, if equilibrium involves work in the period of decision, a lower average consumption rate after retirement than before is consistent with quasi-hyperbolic but not with exponential discounting[27].

Interesting comparisons of comparative statics nature also emerge. Consider an equilibrium in the three-period model in which the agent undersaves in the first period and works in the second. If earnings in the second period (Δ) increase, the period 1 self will save *more*: the extra earnings gives self 2 more incentive to work, lowering the amount of undersaving needed to induce work. Thus, self 1's marginal propensity to consume out of changes in future earning is negative. This could never happen with an exponential discounter.

The complete comparative statics for first-period savings with respect to Δ is illustrated in figure 5 for the three-period model. Savings with and without commitment are shown. For very low levels of Δ it is not worth working, so the agent just saves from her other wealth for retirement. These savings don't depend on Δ, as period 2 income is never realized. In the next region, self 1 would prefer self 2 to work if she could commit her to do it, but, without it, it is better to retire early. The interesting region is the next one. Here, self 1 undersaves to make self 2 work, giving her exactly $s^* = \overline{W}_2$. Since $\overline{W}_2$ increases with Δ, s^* is increasing; furthermore, this is the only region in which s^* is in general not a linear function of Δ. And finally, for high levels of Δ the equilibrium involves work, and it is once again equivalent to the commitment solution. Notice that

[27]It is tempting at first to try to use this as an explanation for the drop in consumption at retirement. There a number of problems, though: first, the drop in consumption occurs at period t=-1 the latest, i.e. *before* retirement. Also, the drop is much too general of a finding for this theory: it happens to almost all groups of people, irrespective of wealth or when they retire.

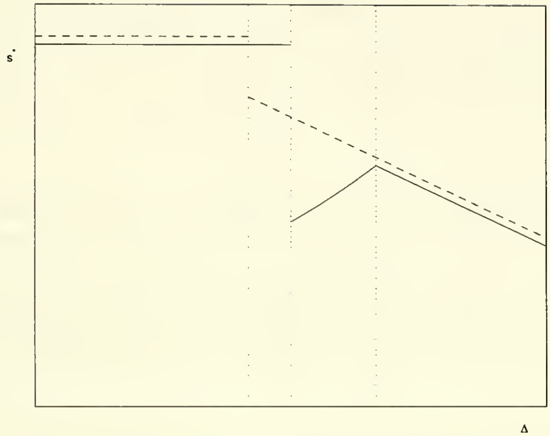

Figure 5: Savings of self 1 with (dashed line) and without (continuous line) commitment as a function of self 2's wage income

Note: the figure is only qualitative; it is not meant to illustrate actual slopes or relative sizes for the regions.

these regions are in exactly the opposite order as in figure 3–higher levels of wealth and higher levels of earnings have opposite incentive effects for retirement.

To check for this effect in practice, one might use the path of consumption to identify people who are undersaving, and then check whether agents with a higher income-to-wealth ratio save a larger proportion of their wealth for the period of decision.

Similarly, if the option to retire in the second period is eliminated in some way, and self 1 would have undersaved before, she will save more. This is again impossible with exponential discounting: there the elimination of a non-chosen alternative doesn't change the optimum. Also, agents who work in the second period in equilibrium but don't undersave, will not change their behavior. Therefore, it is probably econometrically easier to check this effect: it is not necessary to identify undersavers, an exercise that carries with it the backbreaking task of disentangling undersaving from other peculiarities in the consumption path that happen around retirement. It is, however, necessary to identify those who would have worked had the option been available.

On the other hand, if the option to work is eliminated, the comparative statics are uncertain. We have seen that savings could go up or down in this case.

Finally, notice that in the long-horizon equilibrium there is a range of wealth levels where richer people save disproportionately more of their wealth for retirement: as one switches from lower savings and work to retirement, the total wealth that self 0 gets switches from $\overline{W}_0$ to something that is greater than $\overline{W}_0 + \lambda^* \Delta$, a change that is not warranted by the difference in lifetime wealth. Thus,

controlling for income, on average the richer people (who retire early) have higher savings rates. This is 'anecdotally' consistent with the stylized fact that wealthier people appear to be slightly more patient, a finding that can't be explained by exponential discounting without appealing to individual heterogeneity in time preference.

7 Naive agents

As economists, we often assume too much rational capability on the part of humans. Our assumption of full sophistication of the agents is not immune from this criticism. Thus, we study less-than-sophisticated agents in this section. The opposite extreme assumption to sophistication is naivité.

An agent is called naive if each of her intertemporal selves assumes that future selves will make the same consumption and/or retirement decisions as she would. There is no game in this case, and the 'plans' (current decisions and expectations about future decisions) are simply updated each period. For a contrast of sophistication and naivité in the context of quasi-hyperbolic discounting, see [9].

We work in the long-horizon setting. First, let us assume that the agent is naive only about the retirement decision, not consumption. That is, she still plays a Laibson game with respect to consumption, but each self $t < 0$ assumes that self 0 (and others) will make the same retirement decision as she would. This assumption is mostly for analytical convenience, so that the discussion fits more naturally into what we have been doing. But it might also be interesting empirically, because retirement is (mostly) a one-time decision, so people should have less of a chance of learning about their intertemporal conflicts in this area than regarding consumption[28]. The assumption implies that self $-n < -1$ expects to work in period 0 iff

$$V(W) - \beta\delta^n e \geq V(W - \frac{1}{R^n}\Delta), \tag{16}$$

where period $-n$ wealth, W, now includes discounted earnings in period 0. We have the following result:

Theorem 1 *If self $-n < -1$ plans to work in period 0, so does self $-n+1$.*

This theorem is the consequence of two considerations, one specific to quasi-hyperbolic discounting and one not. First, the Euler equation for consumption implies that the marginal utility of wealth today is less than $R\delta$ times the marginal utility tomorrow, so an extra amount of income that is R times as much in the future as today should be worth more than $\frac{1}{\delta}$ times in the future as today. And since in the future the cost will be perceived to be $\frac{1}{\delta}$ times as much, the future self is more likely to want to work. Second, the future self is additionally motivated to want to work as the current self consumes part of the planned income in period 0. The latter argument does not rely on quasi-hyperbolic discounting, while the former one does. The proof of theorem 1

[28]Note that with this assumption consumption decisions in each period are the same as in the commitment case.

takes advantage of both in a tricky way. It could be simplified, but the given form allows for two generalizations.

It is easy to show that if $(\lambda^*\beta + 1 - \lambda^*)^n \leq \beta$, then self 0 will actually work. Also, though the problem seems different on the surface, both the theorem and the conclusions below are exactly the same if the agent is also naive about consumption decisions.

The converse of theorem 1 is not true–if self $-n < -1$ wants to retire early, self $-n+1$ might change her mind.

It is, however, true that if self -1 wants to retire in period 0, self 0 will actually do so. To see this, note that if self 0 were to work, that would be better for self -1 as well, and with optimal consumption it would be better still. Again, the converse of this is not true: it could happen that self -1 plans to work in period 0, but self 0 decides to retire. In this case, lifetime wealth is updated downwards (self -1 believes that period 0 earnings are a part of wealth), so there is a downward jump in the consumption path. In contrast to the sophisticated case, this occurs *exactly* at retirement, as actually observed empirically [5].

8 Conclusion

This paper makes an addition to the classic quasi-hyperbolic discounting savings model. Its technical contributions are minor–most of the analysis is possible with little more than the tools developed by David Laibson. However, the interaction of two decisions, with the one (savings) available as a tool to influence the other (retirement), changes the classic model in a few interesting ways.

One is the possibility of additional undersaving with the eventual consequence of making the self with a choice poor enough so that she will want to work. This undersaving occurs *in addition* to the undersaving that characterizes the equilibrium without a retirement decision. It therefore aggravates an already inefficient outcome, and, not surprisingly, is likely to be bad for all selves.

The other, and perhaps more novel, effect is the possibility of higher saving. Higher saving can occur for two reasons: either because it is too costly in terms of discounted utility to make the deciding self (self 2 in the three-period models, self 0 in the others) work, and thus one would rather finance her retirement, or because self $t \leq -1$ is too eager to work long and it is worth saving more to make her choose early retirement. Unlike undersaving, it is not in general bad for the individual–it can mitigate the overconsumption equilibrium of the classic model. In fact, higher saving seems never to be Pareto-worsening: the later selves, at least, should be happy about getting more savings. For this reason, it might not be as important in practice as the undersaving equilibrium.

Testing for quasi-hyperbolic discounting empirically runs into one major problem: individual marginal propensities to consume, which are crucial in these models, are basically impossible to measure. Thus a direct test of the Euler equation for consumption is practically infeasible. The retirement decision parameter, on the other hand, allows for good indirect tests, based on either the consumption path or comparative statics. It might be interesting to see some of these tests done. Particularly interesting would be 'natural experiments' changing work and retirement opportunities.

The theoretical model would benefit from two major extensions. One is the introduction of more periods when the agent can choose whether to work. We solved a model of this sort without

savings: in each period, the agent can decide whether or not to retire (the retirement decision being final,) and consumption just equals income or benefits. To make it an interesting problem, one has to assume, for example, a benefit profile that increases with the age of retirement. In equilibrium, the agent retires too early: the retirement date is Pareto-dominated by a later retirement date. No such results emerge in our models with savings, but they might if there are more periods of retirement decisions[29]. Also, it would be interesting to see how consumption unfolds during the periods of decision, and how the different retirement dates are distributed among these periods.

The other useful extension would be the investigation of liquidity constraints in this context. They are clearly important in practice, and they change the nature of equilibria with quasi-hyperbolic discounting considerably.

A perplexing aspect of quasi-hyperbolic discounting models is a question that is very hard to answer: why don't people take advantage of annuity-type commitment devices to overcome their undersaving problem? These financial tools are readily available but rarely used. Some modestly satisfactory reasons can be brought up. First, if there is a bequest motive, then, just like in many exponential discounting models, annuities look less attractive than without a bequest motive. Second, the annuities market is quite complicated, and there are good reasons for boundedly rational people not to enter markets they know little about. The latter seems to indicate that as people learn about annuities they will come into broader use. Even if that happens, the commitment is unlikely to be full, leaving at least some room for quasi-hyperbolic discounting effects. In the absence of annuities, there is of course a wide-spread institutional structure that serves as a commitment device for agents happy or unhappy about it: social security. We plan to study the implications of the joint mandates of savings and receipt of social security benefits as a real annuity in a later paper.

A Existence and uniqueness of equilibria

In this section, we outline a proof of the existence of equilibrium for the long-horizon game. It just requires pulling together much of what we have already shown.

For the game after retirement, the existence and uniqueness of the subgame-perfect equilibrium has been established by David Laibson. For earlier periods, we prove the following general theorem:

Lemma 3 *A Markov-perfect subgame-perfect equilibrium exists with the following properties. For $t \leq 0$, the domain $(0, \infty)$ of the consumption rule $c_t(W_t)$ can be divided into finitely many disjoint intervals such that in the interior of each interval,*

1. the eventual period 0 work/retirement decision is the same,

2. the equilibrium consumption schedules $c_s(W_t)$ for $s > t$ are all differentiable in W_t, and

3. self t has a constant marginal propensity to consume;

4. further, at an interval endpoint a, self t is indifferent between following the limit of the two neighboring intervals' consumption rules, and utility is continuous in wealth at a.

[29] The Pareto-improving retirement date is at least two periods later than the equilibrium date t: otherwise self t wouldn't want to retire. Then it is not a major surprise that our models don't generate too early retirement.

Proof. Starting from $t = 0$, use the following backward induction type of construction for finding the equilibrium: given the next self's strategy, maximize utility for self t. If for some wealth self t is indifferent between a number of consumption levels, assign to her the strategy that the earlier self would prefer.

Of course, we have to prove that this construction works and yields an equilibrium with the above properties. We do this by backward induction.

The case is clear for $t = 0$. Now suppose the statement is true for $t = m + 1$. We will prove it for $t = m$.

Suppose W_m is given. For self $m + 1$, let the intervals in question be divided the by points $0 < a_1 < \ldots < a_M$. For any $\epsilon > 0$, self m's maximization problem has a solution if her savings level is restricted to lie in the interval $[a_i + \epsilon, a_{i+1} - \epsilon]$. Since there are only finitely many intervals, a maximum on the union of these intervals and the points $\{a_i\}$ also exists. It is easy to see that as ϵ approaches zero, eventually the maximum doesn't change. For otherwise there would be a point a_i such that as W_{m+1} approaches a_i from one of the sides, self m's utility is greater than at savings level a_i, which contradicts that when indifferent, self $m + 1$ chooses the consumption level self m prefers[30].

This shows that for each wealth level W_m, self m's problem has a solution. Now define $0 = b_0 < b_1 < \ldots < b_N < \infty$ such that for each $i = 0, \ldots, N-1$, if $W_m \in (b_{2i}, b_{2i+1})$, then $W_{m+1} \in (a_i, ai + 1$, and if $W_m \in (b_{2i+1}, b_{2i+2})$, then $W_{m+1} = a_{i+1}$[31].

By definition, point 1 is satisfied for each (b_j, b_{j+1}). It is also clear that for any (b_{2i+1}, b_{2i+2}), points 2 and 3 are satisfied as well. Therefore let us concentrate on the case $W_m \in (b_{2i}, b_{2i+1})$. Since all future consumption schedules are differentiable at $W_{m+1}(W_m)$, the discounted utility of self m as a function of c_m is differentiable at $c_m(W_m)$. Now $c_m(W_m)$ maximizes this utility, so the derivative at that point is zero. Taking the derivative for selves m and $m + 1$ (as in Laibson [6]) and substituting leads to the Euler equation

$$\frac{u'(c_m)}{u'(c_{m+1})} = R\delta(\beta\lambda_{m+1} + 1 - \lambda_{m+1}), \tag{17}$$

where λ_{m+1} is self $m + 1$'s marginal propensity to consume . Then self m's marginal propensity to consume λ_m on (b_{2i+1}, b_{2i+2}) is constant and is given by the equation

$$\frac{\lambda_m}{1 - \lambda_m} = \frac{R\lambda_{m+1}}{[R\delta(\beta\lambda_{m+1} + 1 - \lambda_{m+1})]^{\frac{1}{\rho}}}. \tag{18}$$

(This is just Laibson's recursion for the λs.)

Also, clearly, utility is continuous in wealth at each interval endpoint, otherwise the agent would 'jump' to the other interval at a different place. Finally, we need to show the agent is indifferent

[30]More precisely, there is a sequence $W_{m+1,n}$ approaching a_i from one side such that discounted utility for self m is increasing on that sequence, and the limit of the discounted utilities is more than discounted utility at a_i. But if at wealth level a_i self $m + 1$ consumes $\lim c_{m+1}(W_{m+1,n})$, by point 4 the discounted utility of self m should be the limit of the discounted utilities when leaving savings $W_{m+1,n}$. But this is impossible by construction as we have assumed that when indifferent, self $m + 1$ does what self m prefers.

[31]Of course, some of the intervals (b_j, b_{j+1}) may be empty.

between the limits of the two neighboring consumption rules. Suppose by contradiction that, say, consuming $\lim_{W_t \searrow a} c_t(W_t)$, doesn't yield the limit of the utilities. This could only be because one of the future selves jumped at an interval endpoint. Then self m's utility actually increased, because when indifferent future selves do what self m wants them to (with quasi-hyperbolic discounting, all previous selves want the same thing.) But in this case near a self m's choice of consumption wasn't optimal, a contradiction. □

In terms of outcome, and with the qualification that the first self alive might be indifferent between two consumption levels at finitely many points, this equilibrium seems to be unique up to outcome: if an interval endpoint is a maximum for an earlier self for some wealth level, then the assumption that the self does what the earlier one prefers is necessary (otherwise that self's problem has no solution), and else it is irrelevant.

B Proofs of some claims

To prove lemma 1, we need the following preliminary result.

Lemma 4 *For $t \leq -1$, savings is monotonically increasing in wealth.*

Proof. Suppose by contradiction that $W_t > W_t'$ but that the corresponding savings levels satisfy $W_{t+1} < W_{t+1}'$. Let the consumption levels be c_t and c_t' and denote the continuation utilities from leaving wealth levels W_{t+1} and W_{t+1}' by K and K', respectively. Further, define $c_t'' = W_t - \frac{1}{R}W_{t+1}'$, $c_t''' = W_t' - \frac{1}{R}W_{t+1}$. Then

$$u(c_t'') + K' \leq u(c_t) + K \tag{19}$$

$$u(c_t''') + K \leq u(c_t') + K'. \tag{20}$$

We can add these and eliminate K and K' to get

$$u(c_t'') + u(c_t''') \leq u(c_t) + u(c_t'). \tag{21}$$

But notice that $c_t > c_t'', c_t''' > c_t'$ and $c_t + c_t' = c_t'' + c_t'''$. Since u is concave, the inequality 21 is impossible. This completes the proof. □

Lemma 1 *Suppose that $t \leq 0$ and $W_t > W_t'$. Then it is not possible that self t with wealth W_t behaves so that self 0 eventually works, and with wealth W_t' she behaves so that self 0 eventually retires.*

Proof. We prove by backward induction. The statement is clearly true for $t = 0$.

Suppose the statement is true for $t = m + 1$. We will prove by contradiction that it is true for $t = m$. Suppose it isn't. Then there are wealth levels W_m and W_m' such that $W_m > W_m'$ and with wealth W_m self 0 eventually works, and with wealth W_m' self 0 eventually retires. Since our statement is true for $t = m + 1$, we then need to have $W_{m+1} < W_{m+1}'$. But this is impossible by lemma 4. □

Lemma 2 *Let $\overline{W}_t$ $(t < 0)$ be the level of wealth at which self t is indifferent between behaviors that eventually lead to self 0 working or retiring*[32]*. At this savings level, self $t-1$ strictly prefers self t to choose to eventually make self 0 retire.*

Proof. We again prove by backward induction, although, as the reader will see, the need for that is little more than technical. Let c_t^w, K_t^w and c_t^r, K_t^r be the consumption levels and continuation utilities for self t with wealth level $\overline{W}_t$ in the working and retirement cases, respectively.

Suppose first that $t = -1$. We have $c_{-1}^w > c_{-1}^r$, since otherwise self -1 would have to leave $\overline{W}_0$ for self 0, which would not make him indifferent between working and retiring. Also

$$u(c_{-1}^r) + \beta\delta K_{-1}^r = u(c_{-1}^w) + \beta\delta K_{-1}^w. \tag{22}$$

To see what self -2 would want, we have to compare $\beta\delta u(c_{-1}^r) + \beta\delta^2 K_{-1}^r$ and $\beta\delta u(c_{-1}^w) + \beta\delta^2 K_{-1}^w$. This is easy:

$$\begin{aligned}\beta\delta u(c_{-1}^r) + \beta\delta^2 K_{-1}^r - \beta\delta u(c_{-1}^w) - \beta\delta^2 K_{-1}^w &= \\ \beta\delta(u(c_{-1}^r) - u(c_{-1}^w)) + \beta\delta^2(K_{-1}^r - K_{-1}^w) &= \\ \beta\delta(u(c_{-1}^r) - u(c_{-1}^w)) + \delta(u(c_{-1}^w) - u(c_{-1}^r)) = (\delta - \beta\delta)(u(c_{-1}^w) - u(c_{-1}^r)) &> 0\end{aligned}$$

If the statement is true for t=m+1, then since self m+1 is not indifferent between self m+2 working and retiring at $\overline{W}_{m+2}$, we have $c_{m+1}^w > c_{m+1}^r$. Then the same proof as above works. □

Theorem 1 *If self $-n < -1$ plans to work in period 0, so does self $-n+1$.*

Proof. If self $-n$ plans to work, next period's wealth is $(1-\lambda^*)RW$. Now

$$\begin{aligned}&\left[V((1-\lambda^*)RW) - V\left((1-\lambda^*)RW - \frac{1}{R^{n-1}}\Delta\right)\right]\delta > \\ &> \left[V((1-\lambda^*)RW) - V\left((1-\lambda^*)RW - \frac{1}{R^{n-1}}\Delta\right)\right]\delta(\lambda^*\beta + 1 - \lambda^*)\end{aligned} \tag{23}$$

since $\beta < 1$. From equation 29 in the appendix, this equals

$$\begin{aligned}&\frac{1}{R}\left(\frac{1}{1-\lambda^*}\frac{1}{R}\right)^{-\rho}\left[V((1-\lambda^*)RW) - V\left((1-\lambda^*)RW - \frac{1}{R^{n-1}}\Delta\right)\right] = \\ &\frac{1}{R}\left(\frac{1}{1-\lambda^*}\frac{1}{R}\right)^{-\rho}\int_0^{\frac{1}{R^{n-1}}\Delta} V'((1-\lambda^*)RW - x)dx\end{aligned} \tag{24}$$

Since V is concave, the above is greater than

$$\frac{1}{R}\left(\frac{1}{1-\lambda^*}\frac{1}{R}\right)^{-\rho}\int_0^{\frac{1}{R^{n-1}}\Delta} V'((1-\lambda^*)RW - (1-\lambda^*)x)dx, \tag{25}$$

[32] Though this fact is not necessary here, it should be said that $\overline{W}_t$ exists and is unique. That it exists can be seen from the consideration that both the set of savings levels where eventual early retirement is (weakly) preferred and where eventual late retirement is preferred are closed. This can be proven easily using backward induction. That it is unique follows from a variant of lemma 1 (the proof of which didn't use strict preferences) along with backward induction.

which, since $V(W) = Vu(W)$ for each W, equals

$$\frac{1}{R}\int_0^{\frac{1}{R^{n-1}}\Delta} V'\left(W - \frac{1}{R}x\right) dx = \int_0^{\frac{1}{R^n}\Delta} V'(W-x)dx = V(W) - V\left(W - \frac{1}{R^n}\Delta\right) \tag{26}$$

through a change in variables. Since self $-n$ plans to work, this is greater than or equal to $\beta\delta^n e$. But then $V((1-\lambda^*)RW) - V((1-\lambda^*)RW - \frac{1}{R^{n-1}}\Delta) \geq \beta\delta^{n-1}e$, implying the claim. □

The proof of theorem 1 really only used that

$$\frac{((1-\lambda^*)R)^\rho}{R\delta} < 1, \tag{27}$$

where λ^* is each self's marginal propensity to consume. Even for agents naive about consumption decisions, marginal propensity to consume is equal across periods with a value of

$$\lambda^* = \frac{1 - (\delta R^{1-\rho})^{\frac{1}{\rho}}}{1 - (1 - \beta^{\frac{1}{\rho}})(\delta R^{1-\rho})^{\frac{1}{\rho}}}. \tag{28}$$

Assuming $\delta R^{1-\rho} < 1$, which is necessary for the naive maximization problem to have a solution, it is easily established that the above satisfies inequality 27. The proofs of the other claims in the text carry over quite effortlessly as well.

C A long horizon after retirement

To start, it is necessary to quote one of Laibson's results:

Lemma 5 *As the length of the horizon approaches infinity, the consumption rule converges pointwise to the function $c_t(W_t) = \lambda^* W_t$, where λ^* is the (unique) solution to the non-linear equation*

$$\lambda^* = 1 - (\delta R^{1-\rho})^{\frac{1}{\rho}}[\lambda^*(\beta - 1) + 1]^{\frac{1}{\rho}}. \tag{29}$$

See [6], page 11 for this. Then we can define

$$\begin{aligned} V(W) &= u(\lambda^* W) + \beta\sum_{i=1}^{\infty}\delta^i u(\lambda^* R^i(1-\lambda^*)^i W) = V\frac{W^{1-\rho}}{1-\rho} = Vu(W), \text{ and} \\ D(W) &= \sum_{i=0}^{\infty}\delta^i u(\lambda^* R^i(1-\lambda^*)^i W) = D\frac{W^{1-\rho}}{1-\rho} = Du(W). \end{aligned} \tag{30}$$

These are the hyperbolically and exponentially discounted values of having wealth W[33]. For future reference, note that the constants V and D satisfy $\beta D < V < D$.

[33] Once again, for $u(c) = \ln(c)$, the constants V and D should enter additively. But also, the discussion is altered only trivially by this.

Suppose self 0 can make a retirement decision, but that selves before have to work and selves later have to retire. Self 0 will work iff

$$V(W_0 + \Delta) - e \geq V(W_0). \tag{31}$$

Let $\overline{W}_0$ be the wealth level that satisfies the above with equality. Self -1 will want self 0 to work iff

$$\beta\delta D(W_0 + \Delta) - \beta\delta e \geq \beta\delta D(W_0),$$

or

$$D(W_0 + \Delta) - e \geq D(W_0). \tag{32}$$

Since $D > V$, the inequality 32 will be satisfied more often than 31.

Analogously to section 3, self -1 would like to set $s_r = R(1 - \lambda^*)W_{-1}$ if self 0 doesn't work, and $s_w = R(1 - \lambda^*)W_{-1} - \lambda^*\Delta$ if self 0 works. Now if the better of these options is available (doesn't conflict with self 2's preferred choice), that will be chosen. But it is also possible that self -1 prefers to work ($V(W_{-1} + \frac{1}{R}\Delta) - \beta\delta e > V(W_{-1})$, i.e. $s_c^* = s_w$), but $s_w = R(1 - \lambda^*)W_{-1} - \lambda^*\Delta$ is above $\overline{W}_0$. If the difference is not too large, it is preferable for self -1 to save less than s_c^* and make self 0 work. On the other hand, if the necessary reduction in savings to make self 0 work is a lot, self -1 will rather have self 0 retire with savings s_r.

These are just the cases we have seen before[34]. To see that $s^* < s_c^*$ can actually happen in 'practice,' revisit the hypothetical curves that we have seen in section 3. In the long-horizon case,

$$\begin{aligned} U_r(W_0) &= \beta\delta D(W_0), \text{ and} \\ U_w(W_0) &= \beta\delta D(W_0 + \Delta) - \beta\delta e, \end{aligned} \tag{33}$$

and, as we have already seen, the optimal savings levels on the two curves are $s_w = R(1-\lambda^*)W_{-1} - \lambda^*\Delta$ and $s_r = R(1 - \lambda^*)W_{-1}$. We know that $\overline{W}_0$ is defined by

$$V(\overline{W}_0 + \Delta) - V(\overline{W}_0) = e,$$

so define $\tilde{W}_0$ by

$$D(\tilde{W}_0 + \Delta) - D(\tilde{W}_0) = e.$$

We can easily choose the parameters such that $\tilde{W}_0 > \overline{W}_0 + \lambda^*\Delta$ (choose $\overline{W}_0$ and Δ such that $D(\overline{W}_0 + \lambda^*\Delta + \Delta) - D(\overline{W}_0 + \lambda^*\Delta) > V(\overline{W}_0 + \Delta) - V(\overline{W}_0)$.) Then if W_{-1} is chosen such that s_w is just above $\overline{W}_0$, s_r will be less than $\tilde{W}_0$. Also, $U_w(W_0) > U_r(W_0)$ for $W_0 < \tilde{W}_0$, so even at s_r, the U_w curve dominates U_r. But since s_w is the optimal point on U_w, it clearly dominates the retirement alternative. Finally, if s_w is close enough to $\overline{W}_0$, the loss of utility from lower saving is small, so it will be worth doing it.

[34] And in all cases except when $s^* < s_c^*$, the consumption path looks like that in a simple savings game.

D A more hyperbolic discount structure

The only change we make is to introduce an additional discount parameter $\gamma < 1$ into Laibson's model, which is effective for 2 periods. Thus, self t's discounted utility from consumption is

$$u(c_t) + \beta\gamma\delta u(c_{t+1}) + \beta\gamma^2\delta^2 \sum_{i=0}^{\infty} \delta^i u(c_{t+2+i}). \tag{34}$$

Of course, we have to start from ground zero and solve the savings equilibrium before we can get into questions concerning retirement. The analysis is similar to Laibson's [6], and we will only go through an accelerated version of it.

Backwards induction along with a repeated use of property 2 of CRRA utility functions proves that in each period, consumption is a linear (and thus differentiable) function of wealth. Then in equilibrium self t will choose c_t to satisfy

$$u'(c_t) = \beta\gamma\delta R \frac{\partial c_{t+1}}{\partial W_{t+1}} u'(c_{t+1}) + \beta\gamma^2\delta^2 R^2 \sum_{i=0}^{\infty} R^i\delta^i \frac{\partial c_{t+2+i}}{\partial W_{t+2+i}} \prod_{j=0}^{i}(1 - \frac{\partial c_{t+1+j}}{\partial W_{t+1+j}}) u'(c_{t+2+i}). \tag{35}$$

The similar equation for period t+1 is

$$u'(c_{t+1}) = \beta\gamma\delta R \frac{\partial c_{t+2}}{\partial W_{t+2}} u'(c_{t+2}) + \beta\gamma^2\delta^2 R^2 \sum_{i=0}^{\infty} R^i\delta^i \frac{\partial c_{t+3+i}}{\partial W_{t+3+i}} \prod_{j=0}^{i}(1 - \frac{\partial c_{t+2+j}}{\partial W_{t+2+j}}) u'(c_{t+3+i}). \tag{36}$$

Combining the two we get

$$\begin{aligned} u'(c_t) = \beta\gamma\delta R \frac{\partial c_{t+1}}{\partial W_{t+1}} u'(c_{t+1}) + \beta\gamma^2\delta^2 R^2 \frac{\partial c_{t+2}}{\partial W_{t+2}} (1 - \frac{\partial c_{t+1}}{\partial W_{t+1}}) u'(c_{t+2}) + \\ \delta R(1 - \frac{\partial c_{t+1}}{\partial W_{t+1}})(u'(c_{t+1}) - \beta\gamma\delta R \frac{\partial c_{t+2}}{\partial W_{t+2}} u'(c_{t+2})). \end{aligned}$$

Putting this into a more convenient form leads to the following lemma.

Lemma 6 *The Euler equation for the choice of consumption at time t is*

$$u'(c_t) = \delta R \left(\beta\gamma \frac{\partial c_{t+1}}{\partial W_{t+1}} + (1 - \frac{\partial c_{t+1}}{\partial W_{t+1}})\right) u'(c_{t+1}) - \beta\gamma\delta^2 R^2 \frac{\partial c_{t+2}}{\partial W_{t+2}} (1 - \frac{\partial c_{t+1}}{\partial W_{t+1}}) u'(c_{t+2})(1-\gamma). \tag{37}$$

It is easily seen that for $\gamma = 1$ this reduces to Laibson's Euler equation.

Using this Euler equation, we can show that in a game with horizon T, the consumption rule is $c_t = \lambda_{T-t} W_t$, where the λ's are determined by the recursion

$$\left(\frac{\lambda_{n+2}}{1-\lambda_{n+2}}\right)^{-\rho} = \delta R^{1-\rho}(\beta\gamma\lambda_{n+1} + (1-\lambda_{n+1}))\lambda_{n+1}^{-\rho} - \beta\gamma\delta^2 R^{2(1-\rho)}(1-\gamma)\lambda_n^{1-\rho}(1-\lambda_{n+1})^{1-\rho} \tag{38}$$

with initial value $\lambda_0 = 1$. Though we haven't shown that this converges, it seems to do so: in computer simulations it converged for all values of the parameters that we have tried[35]. The

[35] It must be said, though, that we haven't tried very many values.

existence of a constant marginal propensity to consume far from the end is not technically necessary for what we are going to do, but it is nice to work off a benchmark that has smooth consumption. We will therefore assume that for our parameter values the long-horizon case has a constant marginal propensity to consume of λ^*[36].

As before, we introduce variously discounted value functions. We will need three this time:

$$\begin{aligned} V(W) &= u(\lambda^* W) + \beta\gamma\delta u(\lambda^*(1-\lambda^*)RW) + \beta\gamma^2\delta^2 \sum_{i=2}^{\infty} \delta^{i-2} u(\lambda^* R^i (1-\lambda^*)^i W) \\ Z(W) &= u(\lambda^* W) + \gamma\delta \sum_{i=1}^{\infty} \delta^{i-1} u(\lambda^* R^i (1-\lambda^*)^i W) \\ D(W) &= \sum_{i=0}^{\infty} \delta^i u(\lambda^* R^i (1-\lambda^*)^i W) \end{aligned} \tag{39}$$

It is easy to see that one period before retirement we get the same undersaving possibility as with quasi-hyperbolic discounting. In that case $\overline{W}_{-1}$ is defined by

$$u(\overline{W}_{-1} - \frac{1}{R}\overline{W}_0) + \beta\gamma\delta Z(\overline{W}_0 + \Delta) - \beta\gamma\delta e = V(\overline{W}_{-1}). \tag{40}$$

To see what self -2 wants self -1 to do at this wealth level we want to look at the difference

$$\beta\gamma\delta u(\overline{W}_{-1} - \frac{1}{R}\overline{W}_0) + \beta\gamma^2\delta^2 D(\overline{W}_0 + \Delta) - \beta\gamma^2\delta^2 e - \beta\gamma\delta Z(\overline{W}_{-1}). \tag{41}$$

Using that $V(W) = \beta\gamma Z(W) + (1-\beta\gamma)u(\lambda^* W) + \beta\gamma\delta(1-\beta\gamma)u(\lambda^*(1-\lambda^*)RW)$ and $Z(W) = \gamma D(W) + (1-\gamma)u(\lambda^* W)$, along with equation 40, the above becomes

$$\begin{aligned} &\beta\gamma\delta u(c^w_{-1}) + \beta\gamma^2\delta^2 D(\overline{W}_0 + \Delta) - \beta\gamma^2\delta^2 e - \beta\gamma^2\delta^2 D(\overline{W}_0 + \Delta) - \\ &\beta\gamma\delta^2(1-\gamma)u(c^w_0) + \beta\gamma\delta^2 e + \delta(1-\beta\gamma)u(c^r_{-1}) + \beta\gamma\delta^2(1-\gamma)u(c^r_0), \end{aligned} \tag{42}$$

where the subscripts on c denote the period in question and the superscripts stand for whether retirement or work is chosen. Dividing by δ and regrouping we get

$$-(1-\gamma)[(u(c^w_{-1} + \beta\gamma\delta u(c^w_o) - \beta\gamma\delta e) - (u(c^r_{-1}) + \beta\gamma\delta u(c^r_0))] - \gamma(1-\beta)(u(c^w_{-1}) - u(c^r_{-1})). \tag{43}$$

Using that self -1 is indifferent between working and retiring

$$(1-\gamma)[\beta\gamma^2\delta^2 D((1-\lambda^*)R(\overline{W}_0+\Delta)) - \beta\gamma^2\delta^2 D((1-\lambda^*)^2R^2\overline{W}_{-1})] - \gamma(1-\beta)(u(c^w_{-1}) - u(c^r_{-1})). \tag{44}$$

Dividing by γ, we finally get that the difference 41 has the same sign as

$$\beta\gamma\delta^2(1-\gamma)[\overbrace{D((1-\lambda^*)R(\overline{W}_0+\Delta)) - D((1-\lambda^*)^2R^2\overline{W}_{-1})}^{I}] - (1-\beta)(\overbrace{u(c^w_{-1}) - u(c^r_{-1})}^{II}). \tag{45}$$

[36]In this case, we also get the familiar undersaving outcome.

The second term in this sum is always negative (II is positive), while the other one can be either positive or negative, though it seems it is more often positive (for that we only need $\overline{W}_0 + \Delta > R(1 - \lambda^*)\overline{W}_{-1}$). For $\gamma = 1$, the first term drops out, so the expression is negative, which means that self -2 would want self -1 to retire at this wealth level. This is just what we had before. On the other hand, with $\gamma \neq 1$ and no degeneracy, the first term can be positive, so we don't necessarily get a negative sum. In particular, if the first term is positive and $\beta = 1$, we can only get lower saving (that is, self -2 wants self -1 to work at $\overline{W}_{-1}$.)

In general, both for γ and $1 - \gamma$ close to 0 (both relative to $1 - \beta$), we will get higher saving. This will be clear intuitively as soon as we understand that equation 45 contrasts two conflicts between selves -1 and -2. First, from the point of view of self -2, self -1 discounts too much between periods -1 and 0, as we had before (term II). But also, self -1 discounts too much between periods 0 and 1, that is, she doesn't appreciate the extra consumption from working as much as she should (term I). For β close to 1, the first effect is negligible. For γ close to 0 or 1, the second one is: close to 1 because then the conflict is small, and close to 0 because then the effect is 'too far in the future' (it is very discounted).

This is only a simple extension of the quasi-hyperbolic setup, but it still indicates that lower saving is more likely with hyperbolic discounting. It also captures what appears to be the two most important conflicts between selves -1 and -2 regarding retirement: that from the perspective of self -2, self -1 overweights consumption in period -1 but underweights consumption after period 0 relative to effort in period 0. Their conflicts about consumption in periods after period 0 are likely to be unimportant. Of course, for earlier selves, this discount structure might not be sufficient: it would be interesting to see better approximations. It won't be easy: genuine hyperbolic discount functions generate equilibria that are extremely hard to analyze.

References

[1] George Ainslie. Specious reward: A behavioral theory of impulsiveness and impulse control. *Psychological Bulletin*, pages 463–96, 1975.

[2] George Ainslie and Nick Haslam. Hyperbolic discounting. In George Loewenstein and Jon Elster, editors, *Choice Over Time*, page 59. Russell Sage Foundation, New York, 1992.

[3] Jonathan Baron. *Morality and rational choice.* Dordrecht ; Boston : Kluwer Academic Publishers, 1993.

[4] Douglas B. Bernheim. Do households appreciate their financial vulnerabilities? An analysis of actions, perceptions, and public policy. *Mimeo, Princeton University*, 1994.

[5] Douglas B. Bernheim, Jonathan Skinner, and Steven Weinberg. What accounts for the variation in retirement wealth among U.S. households? *Working Paper*, 1997.

[6] David Laibson. Hyperbolic discount functions, undersaving, and savings policy. *National Bureau of Economic Research Working Paper*, 1996.

[7] David Laibson. Hyperbolic discounting and time preference heterogeneity. *Harvard University Mimeo*, 1997.

[8] W. Mischel and E. Staub. Effects of expectancy on working and waiting for longer rewards. *Journal of Personality and Social Psychology*, pages 625–633, 1965.

[9] Ted O'Donoghue and Matthew Rabin. Doing it now or later. *UC Berkeley Mimeo*, 1996.